A replica approach to the state-dependent synapses neural network

D. Bollé §, G.M. Shim † and B. Van Mol ‡

§‡ Instituut voor Theoretische Fysica, K.U. Leuven, B-3001 Leuven, Belgium

† Department of Physics, Chungnam National University, Yuseong, Taejon 305-764, R.O.Korea

Abstract. The replica method is applied to a neural network model with state-dependent synapses built from those patterns having a correlation with the state of the system greater than a certain threshold. Replica-symmetric and first-step replica-symmetry-breaking results are presented for the storage capacity at zero temperature as a function of this threshold value. A comparison is made with existing results based upon mean-field equations obtained by using a statistical method.

1. INTRODUCTION

It is standard knowledge by now that for the Hopfield model [1] the Hebb rule leads to a critical storage capacity $\alpha_c = 0.138$ [2] while for these type of models with quadratic interaction the optimal storage capacity is $\alpha_c = 2$ [3]. This is due to the fact that the contribution of the noise caused by the weakly correlated patterns becomes larger than the signal of the condensed patterns as α increases. In order to lift this limitation of the Hebb rule a model with state-dependent synapses has been discussed recently [4, 5]. The idea thereby is to introduce a threshold η cutting out of the Hebb rule all patterns whose correlations with the state of the system are smaller than this threshold [5]. These authors propose an energy function for this state-dependent synapse (SDS) model and derive the corresponding fixed-point equations using the so-called heuristically motivated statistical mean-field scheme developed in [6, 7] (see also [8]). In the case of the Hopfield model ($\eta = 0$) this statistical derivation leads to the same results as those derived using a replica symmetric mean-field theory approach [2]. Solving these

§ E-mail: desire.bolle@fys.kuleuven.ac.be

† E-mail: gmshim@nsphys.chungnam.ac.kr

‡ E-mail: bart.vanmol@fys.kuleuven.ac.be

fixed-point equations one finds, for example at zero temperature, an increase in the storage capacity from $\alpha_c = 0.138$ for $\eta = 0$ up to, e.g., $\alpha_c = 0.17$ for $\eta = 1$. A similar effect has been found for the recognition of temporal sequences [9] and for non-monotonic Hopfield models [10].

In this paper we apply the replica method to the zero temperature capacity problem of the SDS model. The aim thereby is twofold. First we want to find out whether the replica symmetric (RS) fixed-point equations derived by the standard replica approach again coincide with the fixed-point equations found with the statistical method. Second since we expect that the RS results are unstable at zero temperature, we want to determine the effects of a first-step replica-symmetry breaking (RSB1) on the capacity.

Somewhat surprisingly, we find that the RS fixed-point equations are different from the results obtained in [5]. This is due to the fact that for $\eta \neq 0$ the assumptions made in [5] that both the overlap with the non-condensed patterns as well as the noise induced by these non-condensed patterns have a Gaussian distribution are incompatible. Keeping only the (standard) assumption that the noise is Gaussian we can improve the calculations using the statistical scheme and show agreement with the replica symmetric approach. Furthermore, in an RSB1 treatment the critical storage capacity increases versus the RS values but up to $\eta = 1$ the increase is relatively small.

The rest of this paper is organized as follows. In section 2 the SDS-model is shortly reviewed. In section 3 the RS approach to this model at zero temperature is outlined and a detailed comparison with the statistical method used in [5] is made. Section 4 contains a discussion of the RSB1 solution. Some concluding remarks are given in section 5.

2. The SDS-model

Consider a network of N neurons which can take the values ± 1 with equal probability. In this network we want to store $p = \alpha N$ patterns $\xi_i^\mu = \pm 1, i = 1, 2, \ldots, N, \mu = 1, 2, \ldots, p$ that are supposed to be independent and identically distributed random variables with probability distribution $\Pr(\xi_i^\mu) = \frac{1}{2}\delta(\xi_i^\mu - 1) + \frac{1}{2}\delta(\xi_i^\mu + 1)$.

Given a configuration $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$, the local field h_i of neuron i is

$$h_i(\boldsymbol{\sigma}) = \sum_{j \neq i} J_{ij} \sigma_j \, , \tag{1}$$

where J_{ij} are the synaptic couplings given by

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{\alpha N} \xi_i^\mu \xi_j^\mu \Theta\left((m^\mu)^2 - \frac{\eta^2}{N} \right) \tag{2}$$

with m^μ the usual overlap order parameters defined by

$$m^\mu \equiv \frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu \sigma_i \tag{3}$$

and $\eta \geq 0$ the threshold parameter. Due to the presence of the step function $\Theta(\cdot)$ only those terms where $(m^\mu)^2 \geq \eta^2/N$ contribute to the synaptic couplings.

The neurons are updated asynchronously according to the well-known Glauber dynamics. We are interested in the zero temperature limit of this dynamics, which can be written as

$$\sigma_i(t+1) = \mathrm{sgn}[h_i(\boldsymbol{\sigma}(t))]\,. \tag{4}$$

For this deterministic dynamics an energy function has been found in ref. [5]

$$H = -\frac{N}{2}\sum_\mu \left((m^\mu)^2 - \frac{\eta^2}{N}\right)\Theta\left((m^\mu)^2 - \frac{\eta^2}{N}\right)\,. \tag{5}$$

3. A replica approach

The energy function (5) has been used in [5] to derive fixed-point equations for the relevant order parameters using the statistical mean-field scheme [6, 7]. The key idea of the latter calculation is to assume that the noise to which the small overlaps with the non-condensed patterns add up is Gaussian.

In the following we apply the replica approach [2, 8] at zero temperature up to first-order breaking and compare our results with the statistical scheme.

3.1. Replica symmetric results

Following the standard approach we calculate the replica-symmetric free energy per neuron for the SDS-model at zero temperature as the limit $\beta \to \infty$, with β the inverse temperature, of its temperature dependent form. For the latter we obtain as a function of the usual order parameters, i.e., the overlap, $m^1 = m$, with the condensed pattern $\mu = 1$, the Edwards-Anderson order parameter, q, and the residual overlap, r, with the non-condensed patterns $\mu \geq 2$,

$$f^{(RS)}(m,r,q,\beta) = f_0^{(RS)}(m,r,q,\beta) + f_\eta^{(RS)}(q,\beta)\,, \tag{6}$$

where

$$f_0^{(RS)}(m,r,q,\beta) = \frac{m^2}{2} + \frac{1}{2}\alpha\beta r(1-q) + \frac{\alpha}{2\beta}\left[\ln(1-\beta(1-q)) - \frac{\beta q}{1-\beta(1-q)}\right]$$
$$-\frac{1}{\beta}\int Dz\,\ln[2\cosh\beta(m+\sqrt{\alpha r}\,z)] \tag{7}$$

$$f_\eta^{(RS)}(q,\beta) = \frac{\alpha}{2}\eta^2 - \frac{\alpha}{\beta}\int Dz \ln\left[1 - \frac{1}{2}\mathrm{erf}\left(\phi_+(\beta)\right) - \frac{1}{2}\mathrm{erf}\left(\phi_-(\beta)\right)\right.$$
$$\left. + \frac{1}{2}\sqrt{1-\beta(1-q)}\,\exp\left(\frac{\beta}{2}(\eta^2 - \frac{qz^2}{1-\beta(1-q)})\right)\left(\mathrm{erf}\left(\phi_+(0)\right) + \mathrm{erf}\left(\phi_-(0)\right)\right)\right] \quad (8)$$

with

$$\phi_\pm(x) = \frac{[1-x(1-q)]\eta \pm \sqrt{q}\,z}{\sqrt{2(1-q)(1-x(1-q))}} \quad (9)$$

and $Dz = dz(2\pi)^{(-1/2)}\exp(-z^2/2)$. In the above $f_0^{(RS)}(\cdot)$ is the free energy corresponding to the Hopfield model ($\eta = 0$) while $f_\eta^{(RS)}(\cdot)$ reflects the effect of the removal of the non-condensed patterns having a small correlation with the state of the system. Furthermore, the fixed-point equations are given by

$$m = \int Dz \tanh\beta(m + \sqrt{\alpha r}\,z) \quad (10)$$
$$q = \int Dz \tanh^2\beta(m + \sqrt{\alpha r}\,z) \quad (11)$$
$$r = \frac{q}{[1-\beta(1-q)]^2} + \frac{2}{\alpha\beta}\frac{\partial}{\partial q} f_\eta^{(RS)}(q,\beta) \quad (12)$$

In the limit $\eta \to 0$ we find back the fixed-point equations for the Hopfield model, as we should. Furthermore, the change in the Hebb rule realized in eq. (2) manifests itself explicitly only in the order parameter r as one would expect. For zero temperature the fixed-point equations (10)–(12) reduce to

$$m = \mathrm{erf}\left(\frac{m}{\sqrt{2\alpha r}}\right) \quad (13)$$
$$r = \frac{1}{(1-c)^2}\left[1 - \mathrm{erf}\left(\sqrt{\frac{1-c}{2}}\eta\right) + \sqrt{\frac{2(1-c)}{\pi}}\eta\exp\left(-\frac{1}{2}(1-c)\eta^2\right)\right] \quad (14)$$
$$c = \lim_{\beta\to\infty}\beta(1-q) = \sqrt{\frac{2}{\pi\alpha r}}\exp\left(-\frac{m^2}{2\alpha r}\right). \quad (15)$$

These results have to be compared with the statistical mean-field scheme of [5]. But first we check the RS stability of this solution (13)-(14) by calculating, as an indication, the entropy of the replica symmetric phase. We find

$$S = S_0 + S_\eta \quad (16)$$

where

$$S_0 = -\frac{\alpha}{2}\left[\ln(1-c) + \frac{c}{1-c}\right] \quad (17)$$

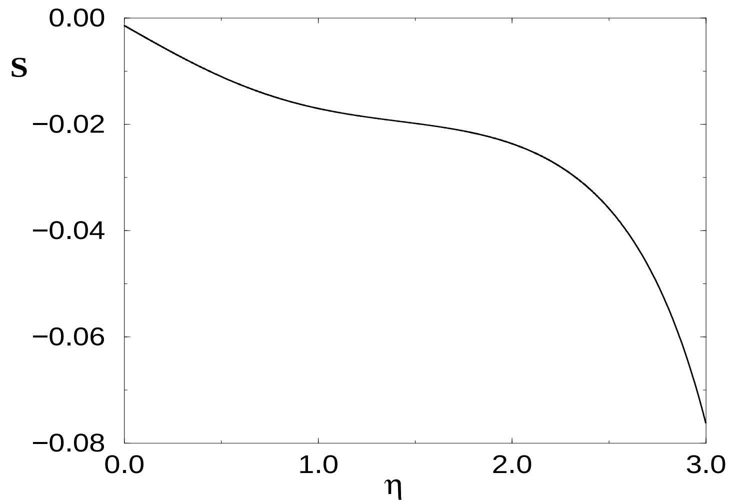

Figure 1. The replica symmetric entropy S at zero temperature as a function of the threshold η.

$$S_\eta = \frac{\alpha}{2}\ln(1-c)\,\mathrm{erf}\left(\sqrt{\frac{1-c}{2}}\eta\right) - \frac{\alpha c}{(1-c)^2}\left[\sqrt{\frac{1-c}{2\pi}}\,\eta\,\exp\left(-\frac{1}{2}(1-c)\eta^2\right) - \frac{c}{2}\mathrm{erf}\left(\sqrt{\frac{1-c}{2}}\eta\right)\right] \quad (18)$$

Again in the limit $\eta \to 0$ the expression (16) reduces to the entropy of the Hopfield model (17) as given, e.g, in [2]. As shown in Fig. 1, the entropy for the SDS-model is negative for all values of η indicating RS-breaking. We remark that for $\eta \to \infty$ the entropy goes to $-\infty$ as $-\eta^{-1}\exp(\eta^2/2)$. To get an idea about the size of the breaking as a function of η a first-order approximation (RSB1) will be performed in section 3.3.

3.2. Comparison with the statistical scheme

Comparing the fixed-point equations (13)-(14) with those of the statistical mean-field scheme derived in [5] (see egs. (23)-(26)) we find that they are different as soon as $\eta \neq 0$. This is illustrated by Fig. 2. It contrasts the situation for the Hopfield model where it is argued [6] that the key assumption in the statistical approach mentioned above – the noise to which the small overlaps with the non-condensed patterns add up is Gaussian – oversimplifies and is most probably responsible for obtaining the results corresponding to the RS approximation. The reason that there is no such correspondence here is that the derivation in [5] not only invokes this key assumption but furthermore supposes that the overlap with the non-condensed patterns themselves, i.e., the $\langle m_N^\mu \rangle, \mu = 2, \ldots, p = \alpha N$ where the brackets $\langle\cdot\rangle$ indicate the thermal average, have

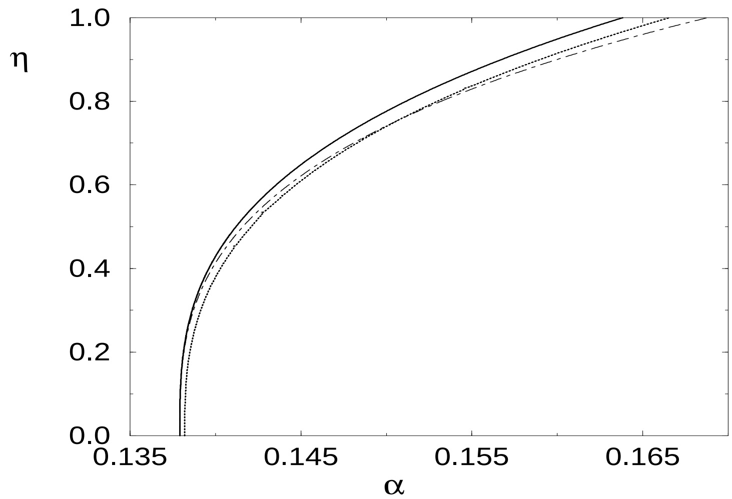

Figure 2. The critical capacity in the RS (full curve) and RSB1 (dotted curve) approximation as a function of the threshold η. For comparison we also show the results of Ref. [5] (dashed-dotted curve).

an identical normal distribution with mean zero and variance σ^2/N. (For convenience we write down explicitly the N-dependence in this subsection). These two assumptions are incompatible for $\eta \neq 0$. Indeed, following closely the derivation in [5, 6] by starting from the mean-field equations for the thermal average of the overlap with a non-condensed pattern, $\langle m_N^\nu \rangle$, and expanding it in a Taylor series to first order we arrive at

$$\left[1 - \beta(1-q_N)\Theta\left(\langle m_N^\nu \rangle^2 - \frac{\eta^2}{N}\right)\right] \langle m_N^\nu \rangle = X_N \tag{19}$$

with

$$X_N = \frac{1}{N}\sum_{j=1}^{N} \xi_j^\nu \xi_j^1 \tanh \beta(m_N + \eta_{N,j}^\nu) \tag{20}$$

$$q_N = \frac{1}{N}\sum_{j=1}^{N} \tanh^2 \beta(m_N + \eta_{N,j}^\nu) \tag{21}$$

$$\eta_{N,j}^\nu = \sum_{\mu \neq 1,\nu} \xi_j^\mu \xi_j^1 \langle m_N^\mu \rangle \Theta\left(\langle m_N^\mu \rangle^2 - \frac{\eta^2}{N}\right) \tag{22}$$

Here $\eta_{N,j}^\nu$ is the noise part and we recall that $\langle m_N^1 \rangle \equiv m_N$ is the overlap of the network with the condensed pattern 1. This expression shows that in the limit $N \to \infty$ the relation between the distributions for the random variables $\langle m^\nu \rangle$ and η_j^ν is no longer simply linear when $\eta \neq 0$. In fact, starting from the key assumption that the noise has a Gaussian distribution, i.e., $\eta^\nu \sim \mathcal{N}(0, \alpha r)$ we find from (20) using the central limit theorem that $\lim_{N\to\infty} \sqrt{N} X_N \sim \mathcal{N}(0, q)$ with q the limit of (21), which is equal to

eq. (11). Furthermore, according to eq. (19) we see that $\langle m_N^\nu \rangle$ is a multi-valued function of the X_N. Employing a standard geometrical Maxwell construction we obtain

$$\sqrt{N}\langle m_N^\nu \rangle = \begin{cases} \frac{\sqrt{N}X_N}{1-\beta(1-q_N)} & \left|\sqrt{N}X_N\right| > \sqrt{1-\beta(1-q_N)}\,\eta \\ \\ \sqrt{N}X_N & \left|\sqrt{N}X_N\right| < \sqrt{1-\beta(1-q_N)}\,\eta \end{cases} \tag{23}$$

The expression (23) clearly shows that the overlaps with the non-condensed patterns are not Gaussian distributed in the limit $N \to \infty$. Using the correct distribution we precisely find eq. (14) for the order parameter r in the zero-temperature limit. Hence, we have shown complete equivalence between the statistical mean-field scheme using only the key assumption that the noise is Gaussian distributed, and the replica symmetric results.

3.3. First-step breaking results

From the observation on the entropy given in section 3.1 we expect RSB effects. In order to get an idea about the size of these effects with growing η we apply first-step RSB. We follow the standard approach (see, e.g., [11]) by introducing the order parameters

$$\begin{array}{rcll} m_\alpha^\mu &=& m^\mu & \forall\alpha = 1,\cdots,n \\ q_{\alpha\gamma} &=& (1-q_1)\delta_{\alpha\gamma} + (q_1-q_0)\epsilon_{\alpha\gamma} + q_0 & \forall\alpha,\gamma = 1,\cdots,n \\ r_{\alpha\gamma} &=& (1-r_1)\delta_{\alpha\gamma} + (r_1-r_0)\epsilon_{\alpha\gamma} + r_0 & \end{array} \tag{24}$$

with n the number of replicas and $\{\epsilon_{\alpha\gamma}\}, \forall\alpha,\gamma = 1,\cdots,n$ a $(n \times n)$ - matrix with elements 1 inside n/k diagonal blocks of size k and 0 outside these blocks.

The free energy per neuron can then be obtained after some tedious calculations

$$\begin{aligned} &f^{(RSB1)}(m,q_0,q_1,r_0,r_1,k,\beta) \\ &\quad = f_0^{(RSB1)}(m,q_0,q_1,r_0,r_1,k,\beta) + f_\eta^{(RSB1)}(q_0,q_1,k,\beta)\,, \end{aligned} \tag{25}$$

where the first term is given by

$$\begin{aligned} f_0^{(RSB1)}&(m,q_0,q_1,r_0,r_1,k,\beta) = \frac{1}{2}m^2 - \frac{\ln 2}{\beta} - \frac{1}{2}\alpha\beta[kq_0r_0 + (1-k)q_1r_1 - r_1] \\ &+ \frac{\alpha}{2\beta}\ln[1-\beta(1-q_1)] \\ &- \frac{1}{k\beta}\int Dz_1 \ln\left\{\int Dz_2 \cosh^k\left[\beta(m + \sqrt{\alpha r_0}\,z_1 + \sqrt{\alpha(r_1-r_0)}\,z_2)\right]\right\} \end{aligned} \tag{26}$$

and the second term reads

$$f_\eta^{(RSB1)}(q_0,q_1,k,\beta) = \frac{\alpha}{2}\eta^2 - \frac{\alpha}{\beta k}\int Dz_1\,\ln\int Dz_2$$

$$\left\{ \exp\left[\frac{\beta z^2}{2(1-\beta(1-q_1))}\right] [1 - \frac{1}{2}\mathrm{erf}\,(\psi_+(\beta)) - \frac{1}{2}\mathrm{erf}\,(\psi_-(\beta))] \right.$$
$$\left. + \frac{1}{2}\sqrt{1-\beta(1-q_1)}\,\exp\left(\frac{\beta}{2}(\eta^2 - \frac{z^2}{1-\beta(1-q_1)})\right)(\mathrm{erf}\,(\psi_+(0)) + \mathrm{erf}\,(\psi_-(0)))\right\}^k \tag{27}$$

with

$$\psi_\pm(x) = \frac{[1-x(1-q_1)]\eta \pm z}{\sqrt{2(1-q_1)(1-x(1-q_1))}}; \quad z = \sqrt{q_0}\, z_1 + \sqrt{q_1 - q_0}\, z_2\,. \tag{28}$$

In the limit $\eta \to 0$ the expression reduces to the Hopfield RSB1 free energy as calculated, e.g, in [12, 13, 14].

Again in the following we are only interested in the zero temperature results. From the limit $\beta \to \infty$ of expression (27) we can obtain the fixed-point equations for the relevant order parameters. Since the way to derive these formula is standard and since their explicit expressions are algebraically complicated we do not write them down.

The zero temperature critical capacity $\alpha_c^{(RSB1)}$ as a function of η is presented in Fig. 2. For $\eta = 0$ we confirm the result $\alpha_c^{(RSB1)} = 0.13819$ found in [13, 14]. For growing η the results for $\alpha_c^{(RSB1)}$ and $\alpha_c^{(RS)}$ start deviating more. In view of the results on the entropy (see Fig. 1) we expect that the difference keeps growing. Since the calculations are very tedious and since in the literature one is mostly interested in values of η smaller than 1 [5, 9, 10] we have plotted results up to $\eta = 1$. For $\eta = 1$, e.g., we find 0.16658 for the RSB1 critical capacity versus 0.16384 for the RS critical capacity. The results of [5] overestimate this value.

4. Concluding remarks

The replica method is applied to an existing neural network model with state-dependent couplings. Only those patterns having a correlation with the state of the system greater than a threshold η contribute to the couplings.

The free energy is obtained and the fixed-point equations are studied at zero temperature. It is shown that the fixed-point equations in the replica symmetric approximation coincide with these found by the so-called heuristically motivated statistical mean-field scheme developed in [6, 7], provided one does not make the additional assumption that the overlap with the non-condensed patterns is Gaussian. This assumption, made in the literature, is totally unnecessary and is even incompatible with the key assumption of the statistical method that the noise induced by the non-condensed patterns is Gaussian.

The critical storage capacity at zero temperature is calculated as a function of the threshold η and compared with the values obtained in the literature on the basis

of the statistical method with the extra Gaussian assumption for the overlaps. Since a calculation of the entropy indicates that replica symmetry is broken at zero temperature for all values of η a first order replica symmetry breaking calculation has been performed. The critical storage capacity increases versus the replica symmetric values but up to $\eta = 1$ the increase is relatively small.

Acknowledgments

This work has been supported in part by the Research Fund of the K.U.Leuven (Grant OT/94/9) and the Korea Science and Engineering Foundation through the SRC program. The authors are indebted to J. Huyghebaert and G. Jongen for constructive discussions. One of us (D.B.) thanks the Belgian National Fund for Scientific Research for financial support.

References

[1] Hopfield J J 1982 *Proc. Natl. Acad. Sci. U.S.A.* **79** 2554
[2] Amit D, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30
[3] Gardner E 1987 *Europhys. Lett.* **4** 481
[4] Marcus I J and Perez P 1990 *Phys. Rev. A* **41** 7013
[5] Zertuche F, Lopez R and Waelbroeck H 1994 *J .Phys A: Math. Gen* **27** 1575
[6] Geszti T 1990 *Physical Models of Neural Networks* (World Scientific, Singapore)
[7] Peretto P 1988 *J. Physique* **49** 711
[8] Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City)
[9] Zertuche F, Lopez-Pena R and Waelbroeck H 1994 *J .Phys A: Math. Gen* **27** 5879
[10] Inoue J-I 1996 *J. Phys A: Math. Gen* **29** 4815
[11] Mézard M, Parisi G and Virasoro M A 1978 *Spin Glass Theory and Beyond* (World Scientific, Singapore)
[12] Crisanti A, Amit D J and Gutfreund H 1986 *Europhys. Lett* **2** 337
[13] Steffan H and Kühn R 1994 *Z. Phys. B* **95** 249
[14] Bollé D and Huyghebaert J 1995 *Phys. Rev. E* **51** 732